DOCUMENT RESUME

ED 441 843 TM 030 878

AUTHOR	Wang, Jianjun
TITLE	Relevance of the Hierarchical Linear Model to TIMSS Data Analyses.
SPONS AGENCY	California State Univ., Bakersfield.
PUB DATE	2000-04-00
NOTE	8p.; Paper presented at the Annual Meeting of the American Educational Research Association (81st, New Orleans, LA, April 24-28, 2000).
PUB TYPE	Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Data Analysis; Elementary Secondary Education; *International Studies; Mathematical Models; *Mathematics Achievement; Sciences
IDENTIFIERS	*Hierarchical Linear Modeling; *Third International Mathematics and Science Study

ABSTRACT

Multilevel international data have been released from the Third International Mathematics and Science Study (TIMSS), providing an opportunity to apply multilevel modeling techniques in educational research. In this paper, TIMSS factors are classified in fixed and random categories according to the project design. Classifying fixed and random factors may help clarify the relevance of hierarchical linear modeling for TIMSS investigations. Relevant methods for the multilevel data analyses are discussed in light of the fundamental relationship between traditional dummy variable regression and hierarchical linear modeling. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 843

Running head: Relevance of HLM to TIMSS

Relevance of the Hierarchical Linear Model to TIMSS Data Analyses

Jianjun Wang
Department of Teacher Education
School of Education
California State University
9001 Stockdale Highway
Bakersfield, CA 93311-1099

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

J Wang

1

TM030878

Paper presented at the 2000 annual meeting of the American Educational Research Association (AERA) in New Orleans, LA. Funding for this presentation is provided in part by the Faculty Teaching and Learning Center (TLC) at CSUB. However, options presented in this article do not necessarily represent positions of the TLC.

BEST COPY AVAILABLE

Relevance of the Hierarchical Linear Model to TIMSS Data Analyses

Abstract

Multilevel international data have been released from the Third International Mathematics and Science Study (TIMSS), which provides an opportunity to apply multilevel modeling techniques in educational research. In this article, TIMSS factors are classified in fixed and random categories according to the project design. Relevant methods for the multilevel data analyses have been discussed in light of the fundamental relationship between the traditional dummy variable regression and the hierarchical linear modeling.

Relevance of the Hierarchical Linear Model to TIMSS Data Analyses

With more than 40 nations participating and five grade levels assessed in two subject areas, the Third International Mathematics and Science Study (TIMSS) is the largest and most comprehensive project in comparative education (Martin & Kelly, 1996). Secondary analyses of the TIMSS data may enrich information on the U.S. condition of education in a cross-nation context (Peak, 1996). Because TIMSS data have a hierarchical structure with students nested in schools and schools nested in countries, some researchers attempted to adopt the hierarchical linear model (HLM) for international comparisons. Based on the eighth grade science data, researchers of the TIMSS International Center reported, "An unconditional, three level analysis of science achievement indicates that the variance can be approportioned 26% within school, 35% between schools, [and] 39% between countries" (Gregory & Shen, 1999, p. 8). Meanwhile, Gregory and Shen (1999) cautioned that the lack of randomization at the country level may not need the variance partitioning in HLM. Since the choice of research methodology inevitably affects results of TIMSS data analyses, the purpose of this study is to examine features of the TIMSS design, and specify proper fixed and random factors within the hierarchical data structure. This investigation may help clarify relevance of HLM for TIMSS investigations.

Fixed and Random Factors in Hierarchical Modeling

Multilevel data can be analyzed by several software packages, such as HLM, MlwiN, and SAS (Wang, 1997). Accordingly, the terminology varies across the different computer software applications. Bryk and Raudenbush (1992) acknowledged in their book for HLM that,

> The models discussed in this book appear in diverse literatures under a variety of titles. In sociological research, they are often referred to as multilevel linear models [cf. Goldstein, 1987; Mason et al., 1983]. In biometric applications, the terms mixed-effects models and random-effects models are common [cf. Elston & Grizzle, 1962; Laird & Ware, 1982]. (p. 3)

At any levels of the hierarchy, factors can be either fixed or random, depending on the research design. Casella and Berger (1990) defined:

> A factor is a fixed factor if all the values of interest are included in the experiment. A factor is a random factor if the values of interest are not included in the experiment and those that are can be considered to be randomly chosen from the values of interest. (p. 529)

Milliken and Johnson (1984) added that "a model is called a mixed or mixed effects model if some of the factors in the treatment structure are fixed effects and some are random effects" (p. 213).

In TIMSS, research interest is confined in around 40 participating nations. Thus, country can be treated as a fixed factor, and the research findings are not designed for generalization over all countries in the world. On the other hand, schools and students are sampled randomly within each nation. Factors at these levels can be treated as random factors, and the sample statistics will be used to represent the condition of education in each nation. Therefore, the hierarchical model contains both fixed and random factors. The HLM computing, as a method of mixed-effects modeling, should be relevant to the TIMSS data analyses.

Dummy Factor Regression

Before TIMSS, the International Association for the Evaluation of Educational Achievement (IEA) sponsored the first and second round of international mathematics and science studies. International comparisons were made using a dummy variable approach in the regression analyses (personal communication with John Keeves, April 13, 1998). For simplicity, a two level model is illustrated in this proposal to articulate relations between the dummy variable regression and the hierarchical linear modeling. According to Bryk and Raudenbush (1992), the hierarchical model can be presented as:

Level 1 (e.g., school) $y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+r_{ij}$

Level 2 (e.g., country) $\beta_{0j}=\gamma_{00}+\gamma_{01}W_j+u_{0j}$
$\beta_{1j}=\gamma_{10}+\gamma_{11}W_j+u_{1j}$ (p. 29)

Besides a random component (r_{ij}) at the school level, the inclusion of country-level variations (e.g., u_{0j} and u_{1j}) implied existence of randomization from the country sampling. By substituting $\beta_{0j}=\gamma_{00}+\gamma_{01}W_j+u_{0j}$ and $\beta_{1j}=\gamma_{10}+\gamma_{11}W_j+u_{1j}$ into $y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+r_{ij}$, the hierarchical model can be further combined in a single equation:

$$y_{ij}=\gamma_{00}+\gamma_{10}X_{ij}+\gamma_{01}W_j+\gamma_{11}X_{ij}W_j+u_{0j}+u_{1j}X_{ij}+r_{ij} \quad \text{(p. 29)}$$

Although a more complex error term ($u_{0j}+u_{1j}X_{ij}+r_{ij}$) has been introduced in the hierarchical model, the lack of random sampling at the country level has led some researchers to question the partitioning of variance at the country level (Gregory & Shen, 1999). Because TIMSS researchers may not be interested in generalizing the results to non-participating nations, the cross-nation differences can described by fixed dummy factors (e.g., W_j) with $u_{0j}=u_{1j}=0$. In this case, the hierarchical model is equivalent to a regression equation with dummy factors:

$$y_{ij}=\gamma_{00}+\gamma_{10}X_{ij}+\gamma_{01}W_j+\gamma_{11}X_{ij}W_j+u_{0j}+u_{1j}X_{ij}+r_{ij}$$

where $r_{ij} \sim N(0, \sigma^2)$.

Bryk and Raudenbush (1992) concurred that "if u_{0j} and u_{1j} were null for every j", the combined HLM model is "equivalent to an OLS [Ordinary Least Square] regression model" (p. 15). Therefore, if fixed factors at the country level are coded by dummy variables, the dummy variable regression is relevant to the hierarchical data analyses in TIMSS.

The differentiation between fixed and random effects can be further delineated by format of the null hypothesis. For instance, due to differences in educational policy and practice, each country has a unique school system. With many schools randomly surveyed in each nation, it would make sense to make inference about the average national performance in TIMSS. Thus, the null hypothesis at the country level is regarding the mean score comparison:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_n$$

for *n* countries participating in TIMSS.

Within each country, since a random sample of schools were selected in TIMSS to represent its condition of education, the school level variables can be specified as random factors affecting student learning experience. Consider a two-level specification between student and school, the HLM can be written as:

Level 1 (e.g., student) $y_{ij}=\beta_{0j}+\beta_{1j}X_{ij}+r_{ij}$
Level 2 (e.g., school) $\beta_{0j}=\gamma_{00}+\gamma_{01}W_j+u_{0j}$
$\beta_{1j}=\gamma_{10}+\gamma_{11}W_j+u_{1j}$

Model in combined form:

$$y_{ij}=\gamma_{00}+\gamma_{10}X_{ij}+\gamma_{01}W_j+\gamma_{11}X_{ij}W_j+u_{0j}+u_{1j}X_{ij}+r_{ij}$$

Unlike the aforementioned levels of country-school hierarchy, u_{0j} and u_{1j} should be retained in the model to reflect the involvement of random school sampling. The corresponding null hypothesis in this case is:

$$H_0: \sigma^2_{school}= 0$$

Thus, statistical null hypotheses may differ across the levels of hierarchical structure, depending on a proper specification of the fixed and random effects. Whether random sampling has been incorporated at the national level does not affect the relevancy of HLM to the TIMSS data analyses. Instead, dummy variable regressions can be adopted along with the random coefficient modeling to specify the mixed structure of fixed and random effects in TIMSS. The clarification of research methodology may facilitate analyses of TIMSS data in a multilevel context.

References

Bryk, A. S. & Raudenbush, S. W. (1992). Hierarchical linear models. Newbury Park, CA: Sage.

Casella, G and Berger, R. L. (1990). Statistical inference. Pacific Grove, CA: Brooks/Cole Publishing.

Gregory, K. & Shen, C. (1999, April). Effective schools: An international perspective using TIMSS science achievement-Grade 8. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Martin, M., & Kelly, D. (1996). Third international mathematics and science study: Technical report. Chestnut Hill, MA: Boston College.

Milliken, G. A. & Johnson, D. E. (1992). Analysis of messy data, Volume 1: Designed Experiments. New York, NY: Chapman and Hall.

Peak, L. (1996). Pursuing excellence (NCES No. 97-198). Washington, DC: U.S. Government Printing Office.

Wang, J. (1997). Using SAS PROC MIXED to demystify hierarchical linear model. Journal of Experimental Education, 66 (1), 84-93.